CA Journal of Cultural Analytics

Measured Applause: Toward a Cultural Analysis of Audio Collections

Tanya Clement and Stephen McLaughlin

05.23.16

Peer-Reviewed By: Jonathan Sterne

Clusters: Sound

Journal ISSN: 2371-4549 DOI: 10.22148/16.002

Applause is a significant cultural marker in recorded performances. In poetry performances, applause can be a means by which an audience can indicate its response to a speaker's performance or to the audience in general; a means for expressing elation and appreciation or, perhaps, dismay; and a way to engage in dialog with a poem itself and affect its mode of meaning making. [1] In the study of a collection of performances, then, applause can serve as a signifier of structures such as the point at which the performance itself has changed from introductory comments to main performance, from single speaker to a question-and-answer period with the audience, or from the end of one poem to the start of the next, but it can also serve, as used here, as a discovery point for considering how a poet interacts with an audience in a particular poetry culture.

Beyond simple annotation and visualization tools or expensive proprietary software, however, open access tools for analyzing aspects of audio such as applause are not widely available for general use by the humanities community. Speech recognition algorithms in projects such as MALACH (Multilingual Access to Large Spoken Archives) are often not built as Web-accessible interfaces for broader audiences. Analysis and visualization software such as PRAAT, which is used by linguists, and Sonic Visualizer, which is often used by music scholars, are desktop tools that typically allow users to focus on one file at a time, making project-sharing difficult for collaborative research and classroom projects. In bioacoustics, researchers use Raven (from the Cornell Lab of Ornithology) and Avisoft (expensive, proprietary software), which perform well with clean data from a single animal. Most of these tools are either not used in multiple domains or with large collections, and none of them do well with the noise or overlapping signals that are often present in historical recordings. As a result of these factors, humanists have few opportunities to use advanced technologies for analyzing large, messy sound archives and sonic cultural markers such as applause remain hidden.

In response to this lack, the School of Information (iSchool) at the University of Texas at Austin (UT) and the Illinois Informatics Institute (I3) at the University of Illinois at Urbana-Champaign (UIUC) have collaborated on the HiPSTAS (High Performance Sound Technologies for Access and Scholarship) project. [2] A primary goal of HiPSTAS is to develop a research environment that uses machine learning and visualization to automate processes for describing unprocessed spoken-word collections of keen interest to humanists. This paper describes how we have developed, as a result of HiPSTAS, a machine learning system to help deal with the challenges that scholars encounter in their attempt to do research with unprocessed audio collections. As a case study, we focus on the acoustic category of applause in the PennSound collection, which includes approximately 36,000 files comprising 6,200 hours of poetry performances and related materials. In doing this analysis, we are able to discern clear differences in rates of applause in reading series that represent different poetry cultures. For those who are interested in implementation, we include an appendix that describes the software used in this paper.

[1] R.S. Gilbert "Joyful Noise: Reflections on Applause," *Southwest Review*, 86(1) (2001): 13-33; C. Goodwin, "Audience diversity, participation and interpretation," *Text: Interdisciplinary Journal for the Study of Discourse*, 6(3) (1986); M. Pfeiler, *Sounds of Poetry: Contemporary American Performance Poets* (Tübingen: Gunter Narr Verlag, 2003).

[2] http://hipstas.org

Use Case: Finding Applause in PennSound Poetry Performances

Why applause?

Humanities scholars have identified audience interactions such as applause as significantly shaping the form and meaning of a public reading. Charles Bernstein refers to literary performance as "both stress test, in which the rhythms are worked out in real time, and trial of the poet's ability to engage listeners." [3] Discussing oral poetry cultures, the French critic Paul Zumthor refers to each audience member as "the coauthor" of a performance. [4] Peter Middleton describes the relationship between audience and poet as a collaboration which "creates an intersubjective network, which can then become an element in the poem itself," [5] and he points to audience interaction as a subject worth further research. [6] For this study, we are working from the premise that applause duration represents a rough index of an audience's engagement with a given reading.

A means for quantifying the presence of applause can lead researchers to consider more in-depth questions such as the relationship between audience response and a poet's performance of the same poem at different venues, as well as the differing responses of audiences at the same venue over the course of a poet's career. We describe example comparisons we made across the PennSound archive below.

Selecting and deploying training examples

For this use case, we ingested approximately 36,000 MP3s (6,200 hours) from PennSound into ARLO. After de-duplication, there were 30,257 files remaining (5374.89 hours). We chose 2,000 files at random, manually examined them for instances of applause, and chose one instance of applause per recording until we had an example training set of 852 three-second tags, including 582 3-second instances of non-applause (3492 0.5-second examples) and 270 3-second instances of applause (1620 0.5-second examples). Optimization for the IBL test went through 100 iterations. As a result of this optimization process, we used the following parameters: 0.5-second spectral resolution; 0.5 damping factor; 0.8 weighting power; 600 Hz minimum frequency; 5000 Hz maximum frequency; 64 spectral bands; spectral sampling rate of 2 (i.e., half-second resolution).

Preliminary Results

We first evaluated our models using cross-validation on the training data. Using the leave-one-out approach, the IBL classifier achieved an overall accuracy of 94.52% with a 0.5 cutoff classification threshold.

	Classified Applause	Classified Non-Applause
True Applause	1509	111
True Non-Applause	169	3323
Accuracy		94.52%

Table 1. IBL Model Training Set Evaluation Confusion Matrix; Average Over 852 Folds

Working with the results produced by the model, we ran tests to understand the optimal smoothing window size and classification cutoff threshold. We created an evaluation set comprising 2,000 files from PennSound known to be full-length public poetry performances. These readings took place between the 1950s and 2010s all over the United States, falling predominantly in the Northeast. They range in length from just a few minutes to over an hour. From these 2,000 readings we selected 10,000 half-second clips at random, manually classifying each as either applause or non-applause. This body of ground truth data allowed us to compare model performance across the two dimensions of our parameter

[3] C. Bernstein, "Reading Voices," In *The Sound of Poetry / The Poetry of Sound,* ed. M. Perloff & C. D. Dworkin (Chicago; London: The University of Chicago Press, 2009), 142-148.

[4] P. Zumthor, *Oral Poetry: An Introduction*, (Minneapolis: University of Minnesota Press, 1990).

[5] P. Middleton, "The Contemporary Poetry Reading," In *Close Listening: Poetry and the Performed Word*, ed. C. Bernstein (New York: Oxford University Press, 1998), 262-299.

[6] P. Middleton, "How to Read a Reading of a Written Poem," *Oral Tradition* 20, no.1, (2005): 7-34.

space: smoothing window size and classification cutoff threshold. In addition to using a standard "flat" rolling average, we also compared the performance of Hann window smoothing.

Because instances of non-applause dramatically outnumber applause in the recordings under study (with applause making up only 1.15% of our ground truth set), overall accuracy is a poor measure of our models' performance. We could, for example, classify every clip as non-applause and claim 98.85% accuracy. The F_1 measure is also ill-suited for mismatched category sizes, as it only considers precision and recall values, disregarding true negatives. We thus used the Matthews correlation coefficient (MCC) as an overall measure of model performance. [7] An MCC value falls between -1 and 1, with 1 representing perfect classification and 0 corresponding to random selection.

After comparing 676 configurations, we found that the optimal approach was using IBL with Hann smoothing over 14 windows (7 seconds) and a threshold of 0.6, achieving an MCC of 0.7606. The accuracy for this configuration was 99.41%.

	Classified Applause	Classified Non-Applause
True Applause	95	25
True Non-Applause	34	9846
Accuracy		99.41%

Table 2. IBL Model Evaluation Set Confusion Matrix Using 14-window Hann Smoothing and 0.6 Classifier Threshold

In our initial exploration of ARLO's IBL classification data, we identified a set of 3,669 public poetry readings, each by a single poet. We removed obviously fragmentary and/or low-quality data by excluding recordings containing less than 2 seconds or more than 100 seconds of reported applause. This left 3,130 readings in our cleaned evaluation set, with a median applause duration of 15.5 seconds and measurements falling in a right-skewed distribution.

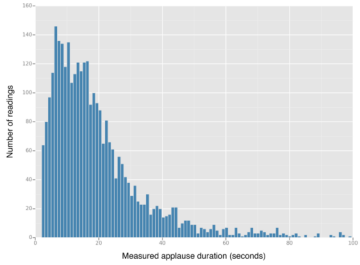

Figure 1. Histogram of Measured Applause Durations for 3,130 Poetry Readings, Binned at 1-Second Level

If we compare the data on recordings by men against recordings by women, as in the plot below, we see that women and men receive similar levels of applause across PennSound. [8] Comparing 1,799 recordings by men and 1,315 recordings by women in our evaluation set, we see that men receive a median of 16.0 seconds of applause versus 14.5 seconds for women, with the difference found to be insignificant using both Student's *t*-test and the nonparametric Mann-Whitney *U* test (discussed below).

	reading length mean	reading length median
Male poets	1914.86	1781.29
Female poets	1551.53	1541.80

Table 3. Mean and Median for Recording Lengths of Male and Female poets

[7] D. M. W. Powers, "Evaluation: From Precision, Recall and F-Measure to ROC, Informedness, Markedness & Correlation," *Journal of Machine Learning Technologies* 2, no.1, (2011): 37-63.

[8] We compiled our gender metadata based on poets' first names, referring to external online sources in cases of ambiguity. This method has precedence as a VIDA Counts methodology (2012). VIDA is a research group committed to advocating for women in the literary arts community. Please see http://www.vidaweb.org/.

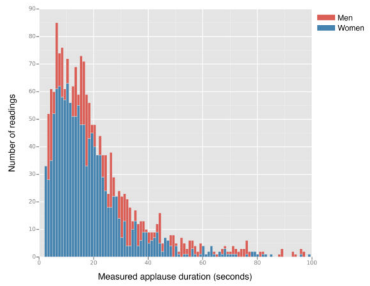

Figure 2. Overlaid Histograms of Measured Applause Duration by Gender for 3,114 Poetry Readings, Binned at 1-Second Level

We then examined applause duration over time, considering measurements for 2,870 recordings between the years 1980 and 2014. The resulting plot demonstrates a stable pattern of audience response over the decades, with the Pearson's correlation coefficient between year and applause duration measured at -0.04 (p=0.01), showing a very small correlation.

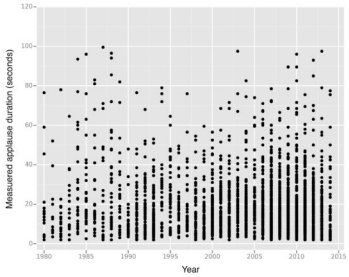

Figure 3. Plot of Measured Applause Durations by Year for 2,870 readings between 1980 and 2014

Next we compared measured durations for readings by six poets, each chosen from the set of performers with ten or more readings in our cleaned examination set. Because we are considering a relatively small number of recordings, and because applause duration is distributed non-normally, we used the Mann-Whitney U test, a nonparametric alternative to Student's t-test which evaluates the null hypothesis that two sets of measurements come from the same population. [9] Table 4 presents pairwise single-tailed tests of applause durations that have been predicted by our IBL classifier. The alternative hypothesis states that the performer in the left column tends to receive more applause than the corresponding one listed in the top row.

	Eileen Myles	Myung Mi Kim	Rodrigo Toscano	Anselm Berri-gan	Bruce An-drews	Elizabeth Willis	Mean Ap-plause Length (sec-onds)	Median Ap-plause Length (seconds)	Mean Recording Length (seconds)	Median Record-ing Length (sec-onds)	No. Read-ings
Myles		0.3208	0.2011	**0.0318**	**0.0135**	**0.0066**	29.65	26.0	2320.76	2272.03	13
Kim	0.7011		0.4102	0.1154	**0.0605**	**0.0300**	27.2	21.0	2019.10	1791.09	10
Toscano	0.8159	0.6189		0.2298	0.1131	0.0732	24.3	20.5	1792.97	2092.18	10
Berrigan	0.9721	0.8977	0.7910		0.4086	0.2028	17.32	17.5	1660.48	1777.16	11
Andrews	0.9876	0.9439	0.894	0.6051		0.2403	16.94	15.25	2389.70	2271.39	24

[9] H. B. Mann & D. R. Whitney, "On a Test of Whether One of Two Random Variables is Stochastically Larger than the Other," *The Annals of Mathematical Statistics*, 18, no.1, (1947): 50-60.

	Eileen Myles	Myung Mi Kim	Rodrigo Toscano	Anselm Berrigan	Bruce Andrews	Elizabeth Willis	Mean Applause Length (seconds)	Median Applause Length (seconds)	Mean Recording Length (seconds)	Median Recording Length (seconds)	No. Readings
Willis	0.9943	0.9742	0.9356	0.8142	0.7700		13.92	12.75	1126.42	1318.19	12

Table 4. *P* values for Pairwise Directional Mann-Whitney *U* Tests Between Six Poets' Measured Applause Durations

Results that are significant at the $p<0.05$ level appear in bold, with the counts and medians of each set of observations provided in the right two columns. It appears, for instance, that recordings of the poet Tom Raworth contain significantly more applause than the others. The recordings of the poet Eileen Myles contain significantly more applause than those by Anselm Berrigan, Bruce Andrews, or Elizabeth Willis. Myung Mi Kim also seems to receive more applause than Andrews or Willis.

Comparing applause durations grouped by city, as in the table below, we find a number of statistically significant differences at the $p<0.05$ level. Measurements from New York and Philadelphia suggest similar applause levels, while both cities' recordings contain significantly more measured applause than Tucson, Oakland, or Boulder. Recordings from New York also contain significantly more applause than those made in Buffalo.

	New York NY	Phila. PA	Buffalo NY	Tucson AZ	Oakland CA	Boulder CO	Mean Applause Length (seconds)	Median Applause Length (seconds)	Mean Recording Length (seconds)	Median Recording Length (seconds)	Median Recording Length (seconds)	No. Readings
New York		0,2255	**0,0347**	**<0.0001**	**0,0065**	**<0.0001**	20,14	17	1757,99		1778,69	1444
Philadelphia	0,7746		0,1196	**<0.0001**	**0,02185**	**<0.0001**	20,11	16	1700,64		1359,06	386
Buffalo	0,9653	0,8805		**<0.0001**	0,1492	**<0.0001**	19,26	15,5	2134,71		2052,05	237
Oakland	0,9935	0,9782	0,8511	**0,0061**		**0,0029**	17,34	13	1505,52		1155,6	85
Boulder	>0.999	>0.999	>0.999	0,8833	0,9971		14,4	8	1427,34		1374,53	74
Tucson	>0.999	>0.999	>0.999		0,9939	0,1173	12,75	10	1425,77		1347,74	97

Table 5. *P* values for Pairwise Directional Mann-Whitney *U* Tests Between Six Cities' Measured Applause Durations (Seconds)

	The Line (NYC)	Chapter & Verse (Philadelphia)	Segue Bowery Poetry Club (NYC)	Belladonna* (NYC)	Mean Applause Length (seconds)	Median Applause Length (seconds)	Mean Recording Length (seconds)	Median Recording Length (seconds)	No. Readings
The Line		**0,016**	**0,0003**	**<0.0001**	27,11	24,5	1663,6	1625,24	63
Chapter & Verse	0,9842		0,3188	**0,0033**	23,54	20,5	1307,22	1232,51	62
Segue Bowery Poetry Club	0,9997	0,6815		**0,0002**	22,42	19,5	1852,77	1848,53	479

	The Line (NYC)	Chapter & Verse (Philadelphia)	Segue Bowery Poetry Club (NYC)	Belladonna (NYC)	Mean Applause Length (seconds)	Median Applause Length (seconds)	Mean Recording Length (seconds)	Median Recording Length (seconds)	No. Readings
Belladonna	>0.9999	0,9968	0,9998		17,96	16,5	1367,13	1420,78	101

Table 6. *P* values for Pairwise Directional Mann-Whitney *U* Tests Between Four Reading Series' Measured Applause Durations

Finally, we compared reading series, which are periodic poetry reading events that happen across a wide variety of venues including bars, coffee shops, bookstores, galleries, and university facilities and often reflect a level of consistency from one reading to the next, not only in terms of format and aesthetics (such as avant-garde or traditional), but also in the set of audience members in attendance. While scholars have examined the underrepresentation of women across poetry publications, [10] there do not appear to be major differences in the rate of applause by gender in our sample of PennSound. We do see differences when comparing cities, which perhaps show evidence of regional variation in communication conventions [11] — but the differences across individual reading series were also more meaningful as units of comparison, since examining applause rates between reading series can provide an opportunity for studying cultural production across a range of different communities.

Table 6 represents a subset of four series from the six in our corpus that included more than 50 complete readings. We excluded two series with over 50 readings — the Left Hand Reading Series and the POG series — because applause at the end of those series' readings is consistently truncated by an editor. Comparing applause rates across these four collections, we see more significant differences than in our tests of individual author pairs. Interpreting results at the $p<0.05$ level as significant, we observe that recordings from The Line Reading Series tend to contain longer total applause durations than those from Chapter & Verse, Segue, or Belladonna. Belladonna readings, in turn, contain significantly less applause in comparison to each of the other three series.

Discussion

Understanding how applause can be mapped to the nature of performances is significant in understanding the cultural context of poetry. A reading series forms a "community of practice" [12] where meaning is being "constituted dialogically through recognition and exchange with an audience of peers, where the poet is not performing to invisible readers or listeners but actively exchanging work with other performers and participants." [13] Indeed, because very different readings can occur at similar venues and include some of the same poets from similar circles, looking at a venue or speaker may not give a consistent picture of the audience dynamics at play at a particular venue or with a particular poet. Ron Silliman notes this phenomena in discussing the different kinds of reactions Robert Glück has received for a poem he wrote about gay bashing: read to a queer audience, he received loud applause for his poem's veracity; read to a university audience, he received quiet appreciation for his form. [14] In contrast, reading series are sites in which poets and their audiences form and maintain particular tastes, conventions, and group temperaments that inform how a poem is shaped and understood and reshaped in that culture. Some series events are serious affairs and result in academic publication; others are community gatherings for fun. Series can be sponsored by a foundation or academic department or run on a shoestring by one or two individuals. Some series provide a venue for early-career writers and those who may feel marginalized by the mainstream

[10] J. Spahr, & S. Young, "Numbers Trouble," *Chicago Review* 53(2/3), (2011):88-111; J. Oggins, "Underrepresentation of Women Writers in Best American Anthologies: The Role of Writing Genre and Editor Gender," *Sex Roles* 71(3-4), (2014): 182-195; VIDA: Women in Literary Arts, March 30, 2016, "The 2015 VIDA Count," retrieved 31/03/2016.

[11] J. Gumperz, "The Speech Community," *International Encyclopedia of the Social Sciences,* Vol. 9, ed. D. L. Sills (New York: Macmillan, 1968, 381-386).

[12] E. Wenger, *Communities of Practice: Learning, Meaning, and Identity* (Cambridge University Press, 1998).

[13] Bernstein, 1998, 63

[14] 24

establishment or are otherwise not formally affiliated with academia. Consequently, series may vary widely in aesthetic focus and degrees of professionalism that are reflected in interactions with the audience.

The Line Reading Series, which receives the highest applause rate in our study, was curated by Lytle Shaw, a professor at New York University, and held at The Drawing Center in New York City between 2000 and 2004. This series hosted well-established, widely acclaimed poets including Jackson Mac Low, Bernadette Mayer, Jennifer Moxley, and Christian Bök who are primarily affiliated with the Language movement and experimental practices such as conceptual writing and Flarf. The second-highest rate of applause occurs in the Chapter & Verse series (run 2008-2012) recordings, which include poets living in Philadelphia (Linh Dinh, Emily Abendroth), many of whom are associated with poetics grad programs at the University of Pennsylvania and Temple University (Brian Teare, Sarah Dowling), alongside a few out-of-town poets formerly affiliated with the SUNY-Buffalo poetics program (Joey Yearous-Algozin, Kristen Gallagher). The third-highest applause rate shown here is associated with the Segue Series, which in the ten years included here (2002-2012), was held at the Bowery Poetry Club in New York's East Village. Founded by Ted Greenwald and Charles Bernstein as the Ear Inn Reading Series in 1978, it has been run by James Sherry's Segue Foundation since 1998, with separate curators for fall, winter, and spring seasons. The series with the least amount of applause is the Belladonna series, run by a feminist avant-garde collective that also runs an independent press. Begun in 1999 and still operating today, Belladonna readings are typically hosted at bookstores, festivals, and other performance venues (such as the Bowery Poetry Club) and readers for the series, almost all of them women, have included Language poets (Carla Harryman, Lyn Hejinian), younger Beat-affiliated writers (Anne Waldman, Leslie Scalapino), and descendants of the New York School tradition (Alice Notley, Erica Kaufman). Like the Chapter & Verse series, Belladonna hosts poets who are both widely acclaimed and affiliated with universities alongside largely unknown artists.

Because meaning at a series reading is constructed with and by a collective audience, [15] the actions or practices that signal these interactions can be pointers to the meaning-making process at play during these events. As communities of practice, reading series include "routines, words, tools, ways of doing things, stories, gestures, symbols, genres, actions" that mark "a history of mutual engagement" that remains "inherently ambiguous," an ambiguity that "is a condition of negotiability and thus a condition for the very possibility of meaning." [16] Applause is part of that ambiguous routine; it is a tool, a way for the audience to express itself and interact with the meaning-making process.

All four of the reading series in our study reflect communities of practice for which audience engagement is particularly important and during which applause happens regularly. The Line Reading Series includes Language movement poets, who create disjunctive poetry that relies on reader and audience participation in the meaning-making process. The Chapter & Verse series (run 2008-2012), which was held in the very small basement of Chapterhouse Café & Gallery in Philadelphia, includes many poets who are emerging into the poetry scene and have relatively few publications. Described on the PennSound website by participants as "expansive and generous about this room that operated outside funding and institutions," the Chapter & Verse series typically draws a less formal, close-knit crowd. The third-highest applause rate shown here is the Segue Series. These performances, held in a bar with a convivial atmosphere, each feature two poets with a younger, less-established poet generally reading first. Many of the headlining poets are associated directly with the Language movement, and openers are in many cases their former students. At these readings, the curators give elaborate, pre-written introductions (roughly five minutes), which are included in these recordings and receive a good deal of applause themselves. Likewise interested in creating a participatory literary experience, [17] Belladonna promotes political engagement and a "feminist literary community among those with a shared (and ever-evolving) poetics." [18]

The communities of all four series are focused on engaging an active audience, but the lower rate of applause in the Belladonna recordings is telling. While the first three series highlight more traditional avant-garde poetry, "where the poet rarely speaks autobiographically and instead presents vocalized artifices of language that might in ordinary discourse be unsayable," [19] the Belladonna series is committed to hosting poets who are first and foremost "political and critical" in a way "that reaches across the boundaries and binaries of literary genre and artistic fields, and that questions the gender binary." [20] In contrast to the traditions of the first three series, which Middleton describes as commonly theatrical and "veiled with a silence about aims," [21] the poets in the Belladonna series are overtly political in their aims and are always

[15] Middleton, "How to Read a Reading of a Written Poem," Bernstein, 1998
[16] Wenger, *Communities of Practice: Learning, Meaning, and Identity.*
[17] R. Levitsky, "Belladonna Books;" *American Book Review* 31(4), (2010): 5.
[18] Belladonna collective. (n.d.), "About Us. Belladonna," retrieved from http://www.belladonnaseries.org/about/.
[19] Middleton, "How to Read a Reading of a Written Poem."
[20] Belladonna collective, n.d.
[21] Middleton, "The Contemporary Poetry Reading," 263

speaking in ways that reflect autobiography in the presence of their situated bodies. Spoken word poet Leah Thorn reflects on the importance of her personal history in the context of feminist poetry: "For me, as a Jew and as a woman, the very act of speaking out, the act of 'coming to voice' [bell hooks], is intrinsically a political one. One of the many ways of ensuring women's powerlessness has been the suppression of voice." [22]

Our initial results, which show the lower rate of applause in the Belladonna recordings, are provocative in that they show how applause can be a barometer for these subtle differences in how these communities negotiate meaning. Corresponding to heterogeneous intentions, poetry performances can provoke or silence applause for a variety of reasons. Given that these are established, ongoing events, it seems safe to assume that each group entails an appreciative, or at least sympathetic, audience. As such, vigorous applause may express delight at lighthearted whimsy as well as deep regard and respect; an appreciation for a turn-of-phrase, idea, or word play; or support for politically charged statements. On the other hand, silence or limited applause is also an appropriate response - it can mean a smaller audience, but in the Belladonna context, it could also mean a lack of support for a political viewpoint, or, conversely, appreciation or an understanding for one. Limited applause can be a profound data point when it reflects a poet's desire to create a particular audience experience, for example. Susan Schultz describes such a potential scenario as the response to Lois Ann Yamanaka's poetry readings. Describing Yamanaka's work as self-estranging poetry about identity construction, Schultz describes how Yamanaka uses Hawaiian pidgin language to "mimi[c] the dominant culture's silencing of pidgin speakers." [23] Likewise, a lower rate of applause in a series like Belladonna can also be telling. As such, this study presents a model for large-scale analysis across poetry recordings (applause vs. no applause) that yields new opportunities for studying how meaning-making processes change across different poetry cultures.

While we produced results that are promising, it is important to note that there are many aspects of the methods we describe above that call for further exploration and testing. For example, recording quality has improved over time, and as a result, some earlier recordings include more noise and thus more false positives. Further, while some of these differences may reflect variations across regional and community conventions, they are also likely influenced by recording and mastering techniques. In addition, as noted, some recordings are truncated at the beginning or end, either unintentionally — a frequent occurrence in the cassette tape era — or intentionally, as in the case of applause cut off by a recording or digitizing engineer. Finally, recordings that are included in the PennSound archive represent curation decisions that favor certain kinds of performers and certain regions of performance over others. Institutional bias does exist in our sample: there are more readings by men in our sample, as well as in PennSound as a whole; and many series that are considered less formal or less academic, such as poetry slams, are also not well-represented in PennSound. All of these factors have an impact on how we should understand these results and necessitate further study

At the same time, a performance is not simply the communication of ideas through words, but entails aural, auditory, and kinesthetic signifiers that consistently go understudied. Developing tools to study these and other elements of performances not only provokes a reconsideration for machine learning processes that we have honed on textual documents, it also expands our understanding of recorded performances as cultural artifacts for study.

Tanya E. Clement, University of Texas at Austin; Stephen Reid McLaughlin, PhD Student at University of Texas at Austin

Appendix:

ARLO (Adaptive Recognition with Layered Optimization) Software
ARLO was developed with UIUC seed funding by David Tcheng and Tony Borries for avian ecologist David Enstrom (2008) to begin exploring the use of machine learning for data analysis in the fields of animal behavior and ecology (http://wiki.arloproject.com/Main_Page). ARLO software was chosen as the software we would develop through HiPSTAS primarily because it extracts basic prosodic features such as pitch, rhythm, and timbre, which humanities scholars have called significant for performing analysis with sound collections. [24]

[22] Middleton, "The Contemporary Poetry Reading," 58.

[23] S. Schultz, "Local Vocals: Hawai'iPidgin Literature, Performance, and Postcoloniality," In *Close Listening: Poetry and the Performed Word*, ed. C. Bernstein, (New York: Oxford University Press), 343-359.

[24] C. Bernstein, *Attack of the Difficult Poems: Essays and Inventions* (Chicago: University Of Chicago Press, 2011); K. Sherwood, "Elaborate Versionings: Characteristics of Emergent Performance in Three Print/Oral/ Aural Poets," In *Oral Tradition* 21 (1), (2006): 119-147; R. Tsur, *What Makes Sound Patterns Expressive?: The Poetic Mode of Speech Perception* (Duke University Press, 1992).

Filter Bank Signal Processing and Spectrogram Generation and Labeling
ARLO analyzes audio by extracting features based on time and frequency information in the form of a spectrogram computed using band-pass filters linked with energy detectors. ARLO spectrograms contain similar information to the more commonly deployed Fast Fourier Transform (FFT)-based spectrograms. However, while the representation of frequency data in FFT is set by sample rate and window size, the frequency filter banks that ARLO uses can be focused on a particular frequency space and optimized for each classification problem. The filter bank method is similar to using an array of tuning forks, each positioned at a separate frequency, an approach that is thought to best mimic the processes of the human ear. [25] With filter banks, users can optimize the trade-off between time and frequency resolutions in the spectrograms [26] by choosing a frequency range and 'damping factor' (or damping ratio), a parameter that determines how long the tuning forks 'ring.' By selecting these features, users can optimize their searches for a given sound pattern.

In ARLO, examples for machine learning are audio events that the user has identified and labeled. Audio events comprise a start and end time, such as a two-second clip, as well as an optional minimum and maximum frequency band to isolate the region of interest. Users label the examples of interest (e.g., "applause" or "barking"). Control parameters are provided for creating spectrogram data according to optimal resolutions for a given problem. Each algorithm described below retrieves the features of the tag according to the user's chosen spectral range and window size (e.g., two frames per second, each 0.5 seconds) from the audio file. We then apply this model to a specified collection of unseen audio files.)

ARLO Machine-Learning Algorithms: Instance-Based Learning
The ARLO instance-based learning (IBL) algorithm [27] searches for the most effective spectrogram representation for a given problem by optimizing all parameters of the spectrogram generation algorithm. Because our goal is to balance accuracy and efficient performance, the IBL algorithm uses an unbiased (weak) optimization method called uniform random search in which each point in the parameter space is equally likely to be evaluated as any other point. While a relatively slow optimization method, it avoids the potential problem of multiple local optima. By default, the parameter space consists of the widest range of possibilities, which can solve a wide range of problems. ARLO searches the parameter space for the best (highest-performing) solution for a given problem. This means ARLO tries many different combinations of spectral extraction parameters and distance weighting powers in an attempt to find a combination of example representation and learning algorithm that works best for the given problem. In this case "best" is a solution that (1) runs in a reasonable amount of time, and (2) has the highest accuracy, which is measured by using leave-one-out cross-validation to simulate performance on unseen examples. Optimization goes through random iterations based on bounds chosen by the experimenter for each parameter (damping factor, minimum frequency, maximum frequency, number of spectral samples, number of spectral bands, distance weighting) until the learning curve demonstrates diminishing returns.

The ARLO IBL algorithm finds matches by taking each known classified example and "sliding" it across new audio files looking for good matches based on a distance metric. Correlation between 64-band spectral vectors is calculated using Pearson's correlation coefficient (PCC), with each pair of corresponding bands considered a single observation. Because PCC falls between -1 and 1, 1 is added to the correlation measure to produce a positive distance value. Classification probability is then calculated using the continuous weighting approach (i.e., kernel density). The class of each member of the training set is weighted according to its distance from the instance to be classified, with weight = 1.0 / (distance) ^ power, where power is determined by optimization. Finally, the average of the weighted training set classes determines prediction probability.

The number of match positions considered per second is adjustable and is set to the spectral sample rate. In addition to simple spectra matching, a user can isolate pitch and volume traces, compute correlations on them, and weight the different feature types when computing the overall match strength. This allows the user to weight spectral uniform random search in which each point in the parameter space is equally likely to be evaluated as any other point. While a relatively slow optimization method, it avoids the potential problem of multiple local optima. By default, the parameter space consists of the widest range of possibilities, which can solve a wide range of problems. ARLO searches the parameter space for the best (highest-performing) solution for a given problem. This means ARLO tries many different combinations of

[25] C. D Salthouse and R. Sarpeshkar, "A Practical Micropower Programmable Bandpass Filter for Use in Bionic Ears," *IEEE Journal Of Solid-State Circuits*, 38(1), (2003): 63-70.

[26] T. D. Rossing and F. R. Moore, *The Science of Sound* (3rd edition.) (San Francisco: Addison-Wesley, 2001).

[27] In order to extend the machine learning capabilities in ARLO, Loretta Auvil and Thomas Redman have also integrated the Weka API, a popular suite of machine-learning tools. Specifically, we have implemented Weka's SVM classifier in order to provide users the opportunity to compare these results against those generated by the IBL algorithm. As opposed to the probability prediction, the SVM algorithm finds a linear hyperplane separating the categories (classes) with the maximal margin in this high-dimensional space and makes a binary classification for a given example. Our implementation of SVM uses n-fold cross-validation with stratification to evaluate the accuracy of the model.

spectral extraction parameters and distance weighting powers in an attempt to find a combination of example representation and learning algorithm that works best for the given problem. In this case "best" is a solution that (1) runs in a reasonable amount of time, and (2) has the highest accuracy, which is measured by using leave-one-out cross-validation to simulate performance on unseen examples. Optimization goes through random iterations based on bounds chosen by the experimenter for each parameter (damping factor, minimum frequency, maximum frequency, number of spectral samples, number of spectral bands, distance weighting) until the learning curve demonstrates diminishing returns.

The ARLO IBL algorithm finds matches by taking each known classified example and "sliding" it across new audio files looking for good matches based on a distance metric. Correlation between 64-band spectral vectors is calculated using Pearson's correlation coefficient (PCC), with each pair of corresponding bands considered a single observation. Because PCC falls between -1 and 1, 1 is added to the correlation measure to produce a positive distance value. Classification probability is then calculated using the continuous weighting approach (i.e., kernel density). The class of each member of the training set is weighted according to its distance from the instance to be classified, with weight = 1.0 / (distance) ^ power, where power is determined by optimization. Finally, the average of the weighted training set classes determines prediction probability.

The number of match positions considered per second is adjustable and is set to the spectral sample rate. In addition to simple spectra matching, a user can isolate pitch and volume traces, compute correlations on them, and weight the different feature types when computing the overall match strength. This allows the user to weight spectral information that might correspond to such aspects as pitch or rhythm. In the IBL algorithm, accuracy is measured using a simulation of the leave-one-out cross-validation prediction process described above.